DOCUMENT RESUME

ED 217 697 FL 012 937

AUTHOR Ricciardi, Joseph S.
TITLE Second Language Tests for Specific Purpose: How Specific?
INSTITUTION Public Service Commission of Canada, Ottawa (Ontario).
PUB DATE Dec 81
NOTE 19p.
JOURNAL CIT Medium; v6 n4 p143-160 Dec 1981

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS English (Second Language); French; Language Proficiency; *Languages for Special Purposes; *Reading Tests; Technical Occupations; *Test Construction; Testing Problems
IDENTIFIERS Quebec (Montreal)

ABSTRACT

This paper discusses certain problems that can arise with respect to the selection of content in the preparation of a test to measure language proficiency for some specific purpose. Forty employees in a defense plant in Montreal, Canada were interviewed in order to specify language needs, to arrive at a precise interpretation of the required reading standard, and to produce appropriate measures of reading proficiency. The main problem involved coming up with a common test that everyone could take and that would contain job-related content. An analysis of the general textual characteristics of work orders and manuals led to the identification of three linguistic skills as the overall testing objectives. Generally these skills involved understanding relationships between parts of sentences and paragraphs, and the ability to identify specific linguistic elements whose characteristics would correspond to the characteristics asked for in a given question. Analysis of the lexical material excluded items that were truly technical. Two conclusions are that the special purposes approach to assessment appears most appropriate when the examinees are as homogeneous as possible, and very accurate definitions of "technical" and "special" are required. An appendix contains examples of item types used. (AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

_MEDIUM, Vol 6 No 4 December 1981
Public Service Commission of Canada

SECOND LANGUAGE TESTS FOR SPECIFIC PURPOSE: HOW SPECIFIC?

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Joseph S. Ricciardi is a project designer-developer in the Official Languages Directorate of the Public Service Commission of Canada's Staffing Branch. He may be reached in Ottawa at (613) 593-5211.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY Michel Lange
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED217697

In the present paper* I would like to discuss briefly certain problems that can at times arise with respect to the selection of content in the preparation of a test to measure language proficiency for some specific purpose, in this instance, reading proficiency for particular job needs. I will illustrate my arguments with some examples drawn from a test which our section has used recently. As yet, the number of persons who have tried this short test is insufficient to permit proper standardization and validation procedures. However, this circumstance is inconsequential for the points I wish to make.

In the Public Service of Canada, second-language requirements of bilingual positions are assigned on the basis of such distinctions as a "general" type of proficiency (with three levels A, B, and C), a specialized level P, and an "instrumental" level R. For the present discussion it is not necessary to provide definitions of the kind of proficiency intended for all of these levels. It will suffice to concentrate on levels R and A of the "general" proficiency type.

Level R is usually assigned, for whatever skill comes into play (e.g., reading or speaking), in situations where some capacity in the second language is required but where the scope and range of the use which a speaker must make of such capacity in his or her job is rather closely circumscribed. As for Level A, the actual proficiency involved is only somewhat higher than that required for Level R, but the communication situations associated with this level are generally less clearly defined.

FL012937

* I would like to thank Gregg Edwards for some useful suggestions for improving the present text. Any shortcomings are naturally mine. Any remarks made here about the system of language levels or connected matters do not necessarily coincide entirely with any official view held by the Public Service Commission of Canada.

With this very cursory look at the levels in question, we may now state the problem which led to the making of the reading test alluded to above. In the Public Service, examinees may contest unsuccessful results of a second-language test if they feel that certain factors may have come into play which adversely influenced their performance on the test. In many instances, such examinees are re-assessed by means of special evaluation instruments and not by the customary Language Knowledge Examination (LKE) which they had undergone in the beginning.

During 1980 the Public Service Commission's Official Languages Directorate received a request, from the Department of National Defence, for the special evaluation of a number of civilian employees in the Department's Workshop 202 at Longue-Pointe, Montreal. The problem, as identified by the Department, was a higher-than-average failure rate among the employees of the Workshop. This high failure rate was alleged to stem from the inappropriateness of the test normally used for the kind of examinee typically found at Workshop 202. Specifically, it was argued that the test's content was not suited to the particular work needs of this unit, given that all the reading materials used in job situations were of a technical nature not at all reflected in the examination the employees had been required to take.

The Language Standards Section of the Official Languages Directorate was therefore given the task of defining the particular work needs in question and of finding appropriate alternate means of assessing the second-language capacity of such employees. In the remainder of this paper we shall focus on the special assessment instrument developed for Reading.

The first step was to undertake an investigation of language-related activities typical of the work milieu of a sample of employees of Workshop 202. These employees represented various trades and they belong to the Public Service's Operational Category; the job areas, in our case, were the following: quality control; engine repairs and refit (ships); welding (general); electrical repair (general); ship refit; radio and radar; carpentry; painting; armaments and maintenance; vehicle repair and maintenance; textiles (e.g., tent repairs); weaponry (general).

Employees were interviewed and observed on the job, and samples of the types of materials that they are normally required to read were collected. In order to verify and round out the information, interviews were also held with the immediate supervisors of some of these employees, as well as with senior military personnel at the management level.[1] The total number of employees interviewed was 40 and the number of work contexts observed was 8. The total number of senior military and supervisory personnel interviewed was 25.

The aim, therefore, was to establish the kinds of reading, listening/speaking tasks required, the purposes for such tasks, and the types of true expectations as to the required level of competence or performance in the second language in relation to these tasks.

Such research permitted a better specification of the second language needs (activities, content of the activities and the level of expectation) for the particular group of employees at Workshop 202. Consequently, it was possible to arrive at a precise interpretation of the "R" standard for Reading and eventually to produce more appropriate measures of reading proficiency.[2] For R in Reading, the standard was defined as follows:

> the ability to recognize key phrases and pick out isolated elements of information from diagrams or drawings and from such abridged texts as work orders or summarized sections of work manuals.

What remained to be done, therefore, was to devise a moderately short reading test for those employees who still had not met the requirements for level R in Reading. The main difficulty connected with this project presented itself at the point where it became necessary to select appropriate reading content for a test. The traditional solution of selecting some general-interest reading material (the typical content of LKE 400 B, the examination the employees to be assessed had already tried and failed) was unacceptable because it represented, in a sense, the type of content which ostensibly the examinees at Workshop 202 were poorly equipped to handle, in terms of their general educational level and prior experience in taking tests of any kind. With such a hypothetical test the problems of content and face validity (i.e., little relevance to job specifics) would still exist.

It was therefore decided with no further ado that the reading material for a test should be sought among the sample of job-related texts that had been gathered during the prior investigation of second-language needs at Workshop 202.

It is worth stressing at this point that our explicit assignment for this project was to produce one reading test.[3] However, even a brief examination of the sample materials gathered earlier was sufficient to indicate that the individual subject matter of each job area was simply too particular to each sector to be used for a common bank of items. The most typical kinds of job-related reading materials were the following, given here more or less in descending order of frequency: work orders, blueprints, and manuals (comprising both diagrams and continuous texts). Thus, the most "true-to-life" test that could have been made would have taken over actual segments of some of these materials. It would have included, besides typical syntactical constructions, some of the job-related terminology easily found in the available materials. In this sense it might have been analogous, in its general approach, to the type of test item one finds illustrated in Carroll (1980), see for example his Table 1 (p. 41), where examinees are expected to be able to deduce certain types of information contained in a "demand and supply schedule." In our case we could have taken, say, a page from a manual dealing with the reconditioning of an engine starter, including a diagram of the starter and a text presenting a summary of procedures for disassembling the starter. Such material would have been ideal in some respects; for example, the fact that all the information required to answer questions on both the text and the diagram was contained on one page would have offered considerable convenience. It would then have been a relatively simple task to test the examinee's ability to read job-related English sufficiently well to meet the requirement of Level R (as defined above).

Unfortunately, such an approach, attractive though it seemed, was not deemed possible. As already suggested above, there was a serious problem of non-transferability of material from one sector to another. Even if the materials available had been so constructed that one did not need to be an expert in the repair of such objects as engine starters in order to be able to "pick out isolated items of information from diagrams," such test items might have been assailable at the very least on the grounds

of poor face validity. Further, even if a test item contained no specifically technical vocabulary, just how would a canvas sewer react to a text about an engine starter? It was therefore clear that we could not respect both conditions at once, that of one common test (which the administrative context required of us) and that of job-related content in a rather specific sense of this term.

Before we describe briefly the solution we finally opted for, one other point must be discussed. As Hutchinson and Waters (1981) argue with respect to the selection of materials for the teaching of subject-specific English, a noticeably different linguistic competence is not necessarily required for each subject in order to study subject-specific English. As the authors demonstrate in their article, it is not necessarily true that, because texts in each subject area can be considered technical in so far as they exhibit certain lexical and syntactical characteristics, persons who must communicate with respect to such subject matter always rely on specifically technical linguistic elements. As the authors rightly maintain, some lexical items and constructions are truly technical while others, though technical in appearance, are only marginally technical: their meaning is easily accessible to non-specialists as well.

With respect to our own particular task of selecting appropriate content for testing purposes, we soon came to conclusions rather similar to the above. Thus, while "bushing," as a component of a starter, could probably be regarded as a technical term in that only someone familiar with starters would know exactly what this term refers to, the meaning of some other starter components, e.g., "contacts," overlaps in part with the meaning they have in other discourse contexts. Moreover, if testing the knowledge of technical vocabulary had been the main assessment objective, a straightforward vocabulary test would probably have been the most recommended testing instrument. Finally, it should be remembered that the employees to be tested were not novices in their respective job domains. They all had many years' experience in their jobs and had already used English for quite some time, even if only intermittently, and were in all probability quite familiar with the truly technical vocabulary of their area. Consequently, for our purpose the expressions "isolated elements of information," given in the definition of Level R, had to refer to something other than technical vocabulary in the strictest sense.

An analysis of the general textual characteristics of the main types of texts available, i.e., work orders and manuals, was undertaken in order to identify the kinds of linguistic operations to be tested and the lexical as well as structural content around which items should be constructed.

Our analysis led us to the identification of the following linguistic skills as the overall testing objectives:

(a) Understanding relationships between segments of a sentence, or two adjacent sentences, by recognizing grammatical cohesion devices and/or such semantic relationships as synonymy, hyponymy, antithesis, or incompatibility at the sentence level.

(b) Understanding relationships between parts of a paragraph by recognizing:
    (i) lexical or grammatical cohesion within a paragraph;
    (ii) the most suitable lexical item or phrase required as a component of grammatically correct and semantically appropriate textual segment within a short paragraph.

(c) Being able to identify, in a passage, specific linguistic elements with characteristics which correspond to those asked for in a given question.

The analysis of the lexical material was predicated on the exclusion of items that were truly technical or in any case very peculiar to a given job type (cf. our discussion above). In other words, an initial distinction was made between "general" and "specific" items, the latter being usually rejected. The following types of selections were made:

(a) Substantives:
    (i) "general" substantives designating objects;
    (ii) "general" substantives designating processes or states;

(b) Verbs:
    (i) "general" verbs designating operations or processes;

(ii) a few "specific" verbs designating processes or operations peculiar to a given trade but nevertheless rather well known by people in general

(c) Other categories:

(i) adjectives and adverbials of the "general" variety;

(ii) some conjunctions, prepositions, and prepositional phrases.

To illustrate, some of the "general" substantives retained are: container, pressure, panel, assembly, adjustment, inspection, etc.; some verbs: loosen, install, replace, measure, etc.

In order to fulfil our objectives by using the kinds of notional/lexical content outlined above, the following types of test items were constructed:

(a) some paragraphs, each of which represents a self-contained text, on which examinees had to answer specific and well-circumscribed questions;

(b) items in which a variety of paraphrase relationships had to be recognized;

(c) a variety of short texts with blanks, as in Cloze tests, in which the blanks had to be filled in; however, unlike the conventional Cloze test, each test question was accompanied by a list of correct elements, in scrambled order, which were to fit into the blanks (the blanks were situated on the basis of the specific lexical and structural relationships we wanted to test, rather than on the basis of every-nth-word deletion as in Cloze tests).

Some illustrations of the test items used are provided in the appendix to this paper.

The test was eventually administered to a number of employees among those for whom special evaluations had been requested (the testing sessions also involved specially adapted

oral proficiency interviews for the assessment of Speaking, an aspect of the problem that we have chosen not to discuss in the present paper). Although the carrying out of special assessments for Level R was the primary aim of our work, the sample of examinees also included a small number of employees who had to meet Level A requirements. Since on LKE 400 B, the test already attempted by these examinees, the cut-off score for Level A is moderately higher than that for Level R, a minimum passing score for Level A was established for the assessment of those who had to meet this latter level.[4]

Several examinees who had previously failed when assessed by LKE 400 B passed the special assessment. However, several who had failed 400 B likewise failed the special test. Since a rather small number of persons actually came to the special assessment, very few, if any, specific conclusions can for the moment be drawn from such results. Perhaps we may merely speculate, on the one hand, that for some examinees the format and content of LKE 400 B was indeed the real obstacle; on the other hand, some of those who failed the special assessment may simply be persons too easily intimidated by tests in general to perform well on any test. Moreover, it is possible that the original assertions made as to the inadequacy of LKE 400 B for the typical employee of Workshop 202 were somewhat overstated: the proficiency level in the second language of some of the examinees may simply be lower than non-specialists perceive it to be. In this case the affirmations originally made on behalf of such examinees, i.e. that they have no trouble whatever communicating on the job, may themselves be out of focus in the sense that, after many years of practical experience in their jobs, such employees can probably fare very well precisely because they rarely have to read anything else other than very familiar and very similar work orders or blueprints.

The truly instructive aspects of this project lie not in the statistics but in what was learned by attempting it. Following our experience with the development of the initial version of this test for a specific purpose, we may conclude the following:

(a) The "special purposes" approach to the assessment of second-language proficiency appears most appropriate and effective when the examinees represent as homogeneous a group as possible with respect to various factors such as: the general level of education of the examinees; the ideational and

propositional content of the texts used in job situations (in the case of the written mode); and the various linguistic operations (or "skills," to use the term from Munby 1978) that must be assessed. The more particular are the premisses for tests, with reference to the above kinds of factors, the less can one type of test devised for one group of examinees be employed for a different group of examinees. This principle is well illustrated by our own case, where a highly selective approach to the task of providing "job-related" content would have indeed given us test items resembling very closely the subject matter of each group of examinees (i.e., corresponding to all of the occupational groups involved); however, such an approach would also have yielded test items with content validity for only a particular type of examinee. While such a solution would have been acceptable on a strictly theoretical basis, at a practical level a judicious selection of content and skills had to be made in order to arrive at a set of elements that could fairly be regarded as appropriate for the vast majority of the examinees in question. Essentially, our solution was to make special uses of not necessarily very special elements.[5]

(b) From (a) above it follows that very accurate definitions of notions such as "technical" and "special" are required for any project of the type described in the present paper. Such definitions concern both the linguistic skills to be assessed and the linguistic content to which the skills are to be applied. Our experience showed us that by careful scrutiny of the job-related reading materials available to us it was possible to arrive at definitions of "job-related" and "special purpose" that generally fit the examinee population while at the same time excluding elements that by their excessive specificity would have hindered our efforts rather than promoting them. While such an approach was the result of a particular administrative contraint applied in our situation, it also has implications for special purposes testing in general. The mere fact that certain types of lexical elements or linguistic constructions occur in texts which by their format, by their typical propositional content and rhetorical devices, and by their overall syntactic regularities are to be considered technical texts, does not necessarily render such lexical elements or constructions "technical" and therefore "special." What we eventually retained for a measure of proficiency was content of a rather everyday kind. Even the linguistic skills that formed part of our testing objective can probably be considered to be necessary for the intelligent

reading of virtually any king of job-related text.[6] Ultimately, the solution adopted shows that to test for specific purposes it is not always necessary to have recourse to content that duplicates unswervingly the terminology specific to a single job area.

FOOTNOTES

1. The approach to the investigation of language need referred to here in necessarily very concise terms is similar in various respects to that discussed in Ricciardi and Edwards (1982).

2. The description of the project up to and including the definition of Level R for employees of Workshop 202 is in part an adaptation of the description provided in Edwards (1981).

3. Here it must be clearly understood that the requirement for a single test represents an administrative constraint within which the project team had to operate, rather than a technical preference on our part. In effect, as we explain in the main body of this text, if circumstances had permitted, a short test for each main job area could have been constructed, especially since appropriate reading materials for most of these areas were available.

4. As we pointed out in our introduction, we do not yet have a sufficient number of scores to render any standardization or validation procedures worthwhile. However, this test may be used again, and thus a sufficient number of observations may become available. For the purpose of the special assessment granted to the employees of Workshop 202, the passing marks for Levels R and A were established in the following manner. A certain number of sub-tests wer given to all examinees. Those examinees who had to attain Level A were given some additional sub-tests. The sub-tests in each group, i.e., the common group and the Level A group, were ranked as to difficulty. Minimum passing scores were assigned to each sub-test in a weighted fashion, i.e., the minimum acceptable score was progressively lowered as the overall difficulty of each sub-test increased. The degree to which a given sub-test involved linguistic capacities felt to be very central to the most typical reading tasks identified earlier also influenced the passing mark set for that sub-test. Examinees who had to pass Level R were required to reach the minimum score on the common group of sub-tests. Level A examinees had to obtain the minimum score on the common group of sub-tests plus the minimum score on the Level A sub-tests. Since far fewer examinees attempted the test than were actually scheduled to

take it, the above was felt to be an appropriate procedure, given the purpose of the test and the particular administrative circumstances in which it had been devised. While the expedient described above does not necessarily correspond to the most conventional procedures for the statistical analysis of test results, it must be emphasized that the official purpose of the test, as far as the Official Languages Directorate was concerned, was simply to ascertain whether the examinees possessed a basic reading proficiency in spite of the fact that they had failed the regular, standardized test. The unofficial, but far more interesting purpose was to experiment with some test items that could form the basis for a possible future standardized test. The customary statistical analyses will therefore be carried out once a more sizable sample of examinees has attempted these experimental items.

5. The papers by Dariya Taylor and Dorothy Roberge, in the present number of Medium, offer good examples of nearly ideal circumstances for the development of tests intended to assess second-language proficiency for a special purpose, since all the examinees are, in one case, graduate students in the History Department of the University of Ottawa and, in the other case, engineering students. In such circumstances each group of examinees would very likely have in common all the important variables that apply in each case. This kind of situation probably presents itself more easily in an academic environment than in the type of occupational context described in the present paper.

6. It is interesting to note in this connection that the books typically intended to aid learners in improving their proficiency in English in their own disciplines subdivide subject matter strictly according to the content typical of each discipline (e.g., the "English in Focus" series published by Oxford University Press). Such a subdivision is of course very appropriate for the purpose. However, as far as linguistic operations are concerned, it is unlikely that they differ greatly from one discipline to another. If it is useful for sociologists to have the linguistic means necessary for drawing comparisons between two sets of facts, it is surely useful for virtually everyone else in other disciplines as well. What probably really distinguishes disci-

plines is, on the other hand, truly particular terminology (unique vocabulary or ordinary vocabulary with peculiar meanings) and, on the other hand, the linguistic structures that occur by virtue of the conventions that characterize the typical modes and channels of each discipline or occupation (e.g., a report of a chemistry experiment vs. a corporation's earnings statement).

REFERENCES

CARROLL, Brendan J. 1980. Testing Communicative Performance. Oxford: Pergamon Press.

EDWARDS, Gregg 1981. Definition of Second Language Needs and Development of Corresponding Measures of Second Language Competence for Operational Category Employees at the Department of National Defence, Longue-Pointe, Montréal: Level "R". Mimeo.

HUTCHINSON, T. and A. WATERS 1981. Performance and Competence in English for Specific Purposes. Applied Linguistics 2: 56-69.

MUNBY, John 1978. Communicative Syllabus Design. Cambridge: University Press.

RICCIARDI, Joseph S. and Gregg EDWARDS 1982. Investigating Occupational Communication Needs in the Perspective of Communicative Performance Testing. T.E.S.L. Talk 13/1:55-74.

APPENDIX

Examples of Item Types Used
for the Test of Reading Proficiency

Première Partie

Cette première partie comprend un texte* ainsi que deux séries de questions. Il vous faudra tout d'abord lire le texte et ensuite répondre aux questions, en suivant les directives données pour chaque série de questions.

EXEMPLES:

(première série)

Dans le texte A, le sens du mot anglais "manufacturer" est très proche du sens de quel mot ci-dessous?

a. mechanic b. clerk c. inspector d. builder

La réponse correcte est (d) parce que parmi les quatre choix offerts, seul le mot "builder" a un sens très proche de celui du mot "manufacturer" (les trois autres mots ayant un sens trop différent ou trop particulier).

(deuxième série)

Trouvez dans le texte A le mot anglais qui signifie "métier", et inscrivez-le dans l'espace ci-dessous:

trade

_____

La bonne réponse est le mot "trade".

* The actual text on which questions are asked could not be included as part of the illustration since the test may be used again in the future.

Deuxième Partie

Cette partie est composée de plusieurs items dont chacun comprend trois énoncés: un énoncé principal (P) suivi de deux énoncés identifiés (a) et (b). Vous devez comparer les énoncés (a) et (b) à l'énoncé (P) et cocher parmi les énoncés (a) et (b) celui qui transmet le message qui se rapproche le plus du message de l'énoncé (P).

EXEMPLE:

(P) All of the mechanical parts of this equipment should be disassembled.

(a) A few parts of this equipment may be disassembled.

(b) In this equipment, the mechanical parts require disassembly.

La bonne réponse est (b) parce que le sens de (b) est plus proche de celui de l'énoncé (P). C'est-à-dire que dans l'énoncé (P), on affirme que toutes les pièces mécaniques doivent être démontées et cette affirmation est reprise par l'énoncé (b); mais l'énoncé (a) ne fait que mentionner "quelques pièces". L'énoncé (b) est donc plus proche de l'énoncé (P).

Troisième Partie

Cette partie comprend un texte représentant des directives de travail. Dans ce texte, on a supprimé une série de mots ou expressions en les remplaçant par des espaces. Ensuite on a répertorié et numéroté, à la fin de ce texte, tous les mots et expressions supprimés. On a également brouillé l'ordre primitif dans lequel se trouvaient ces mots dans le texte.

Il vous faudra relire le texte et, pour chacun des espaces, repérer le mot ou l'expression qui, une fois replacée dans l'espace approprié, servirait à compléter le sens du texte. Lorsque vous aurez identifié le mot ou l'expression juste, il vous restera simplement à en inscrire le numéro dans l'espace visé. Vous aurez ainsi rétabli l'ordre correct des expressions dans le texte.

Vous trouverez ci-dessous un exemple. Lisez-le attentivement.

EXEMPLE

Closely inspect all _____ components of the units. Check for breaks, licks, dents, loose or missing hardware, or other_____.

Check the coils, tubing and tubing_____for apparent leaks and_____.

1. Corrosion
2. Mechanical
3. Defects
4. Connections

Correctement reconstitué, le texte se présente sous la forme suivante:

Closely inspect all mechanical (2) components of the units. Check for breaks, leaks, dents, loose or missing hardware, or other defects (3).

Check the coils, tubing and tubing connections (4) for apparent leaks and corrosion (1).